



英作文支援のための冠詞誤り訂正及びその根拠の提示

著者	梅澤 次郎
学位授与機関	Tohoku University
URL	http://hdl.handle.net/10097/55501

修士論文

英作文支援のための冠詞誤り訂正及びその根拠の提示

梅澤 次郎

2013 年 2 月 13 日

東北大学 大学院
情報科学研究科 システム情報科学専攻

本論文は東北大学大学院情報科学研究科に
修士(情報科学) 授与の要件として提出した修士論文である。

梅澤 次郎

審査委員：

乾 健太郎 教授 (主査)

加藤 寧 教授

伊藤 彰則 教授

岡崎 直観 准教授

英作文支援のための冠詞誤り訂正及びその根拠の提示*

梅澤 次郎

内容梗概

英語における冠詞の使用方法を習得することは、冠詞の概念を持たない言語を母語とする非母語話者にとって、非常に難しいことが知られている。多くの研究者達により冠詞誤り訂正の研究は行われてきたが、完全な訂正は達成されていない。また、これらの既存手法では、正しい冠詞の使用方法を知ることができないため、冠詞の使用方法を学びたい学習者にとって十分な情報が提供されていない。本論文では、冠詞誤り訂正時にその訂正根拠を提示する手法を提案するとともに、訂正根拠を提示することで学習者の冠詞選択能力が向上することを確認した。

キーワード

自然言語処理，英文誤り訂正，冠詞誤り訂正，誤り訂正根拠提示，英作文支援

*東北大学 大学院 情報科学研究科システム情報科学専攻 修士論文, B1IM2008, 2013 年 2 月 13 日.

Presenting the Evidence for the Automatic Correction of English Article Usage*

Jiro Umezawa

Abstract

Mastering proper article usage, especially in the English language, has been known to be an extreme challenge for non-native speakers whose L1 languages do not have any concept of articles. Likewise, the development of correction methods for article usage has posed a challenge for researchers, as current methods do not perfectly correct the articles. In addition, proper article usage is not taught by these methods; therefore, they are not useful for those wishing to learn a language with article usage. In this paper, we discuss the necessity for presenting the evidence for corrections of English article usage. We demonstrate the effectiveness of this approach in order to improve the writing skill of English learners.

Keywords:

Natural Language Processing, Grammatical Error Correction, Article Error Correction, Presenting the Evidence for Error Correction, Writing Support

*Master's Thesis, Department of System Information Sciences, Graduate School of Information Sciences, Tohoku University, B1IM2008, February 13, 2013.

目次

1	はじめに	1
2	関連研究	4
2.1	英作文における誤り	4
2.1.1	冠詞誤り	4
2.2	英文の自動誤り訂正	4
2.2.1	誤り訂正手法に関する既存研究	5
2.2.2	訂正根拠提示に関連する既存研究	6
3	本研究で用いる冠詞誤り訂正手法及び訂正根拠提示手法	7
3.1	課題設計	7
3.2	冠詞誤り訂正モデル	8
3.3	冠詞誤り訂正根拠情報	14
3.3.1	訂正モデルによる冠詞予測の信頼度の提示	14
3.3.2	訂正モデルによる冠詞予測の理由の提示	15
3.3.3	冠詞予測理由に基づく例文の提示	16
4	評価実験	18
4.1	実験に使用したデータセット	18
4.2	実験に使用した分類器の性能の評価	19
4.3	被験者による評価	20
4.4	考察	24
5	おわりに	31
5.1	まとめ	31
5.2	今後の方針	31
	謝辞	32

図 目 次

1	訂正根拠提示システムの概要	8
2	分類器の信頼度スコアと正解率との関連	22
3	構築したシステムの出力例	23
4	根拠情報の有効性	26
5	システムの予測は正しいが、最終的に被験者が冠詞選択を誤った 事例 (1)	30
6	システムの予測は正しいが、最終的に被験者が冠詞選択を誤った 事例 (2)	30

表 目 次

1	Cambridge Learner Corpus における誤りの割合 (スペル誤りを除く)	2
2	本研究で使用した素性一覧	10
3	既存研究の正しい英文上での冠詞予測性能	19
4	各被験者の冠詞予測正解率	24
5	根拠情報の有効性	26
6	根拠提示無しから根拠提示有りでの予測分布の変化	27
7	被験者が誤ってしまった事例の分類	28

1 はじめに

近年、英語非母語話者が英作文をする機会が増加しており、非母語話者の英作文を支援する需要が高まっている。非母語話者の英作文では様々な誤りが見受けられるが、特に、非母語話者は冠詞や前置詞の使用方法についての誤りが多い。例えば、表1に示すように、英語非母語話者による英作文により構成されている Cambridge Learner Corpus [1] における主な誤りは、スペル誤りを除くと、前置詞誤り (13.4%)、限定詞誤り (11.7%) である [2]。限定詞の誤りのほとんどは冠詞の誤りであるため、これらの誤りは実質的には冠詞の誤りであると考えることができる。

このような結果を受けて、前置詞誤りや冠詞誤りを自動で検出・訂正する様々な手法が研究されてきた。これらの研究によって誤り訂正性能は着実に改善されてきてはいるものの、その訂正性能はまだ完全なものではない。例えば、評価型ワークショップ Helping Our Own 2012 Shared Task においてトップの性能を示した Dahlmeier ら [3] の冠詞誤り訂正システムでさえ、学習者の英作文に対する冠詞誤り訂正性能は 31.88% の再現率と 62.93% の精度である。

英文の誤り訂正に関する既存研究の問題点として、ほとんどの既存研究では冠詞や前置詞の訂正の根拠までは示さないため、学習者は訂正情報として提示された冠詞や前置詞と自らがもともと選択していた単語のどちらかを、誤った理由が分からない状態で選択しなければならないという問題がある。それゆえ、誤り訂正システムの実際の使用を想定した場合、学習者の英作文を支援するためには、なぜ誤り訂正システムによって提示された訂正情報が、もともと書かれていた語よりも正しそうであるかを示す根拠情報 (e.g. 対象名詞が複数形の場合、不定冠詞 *a* は付与できない) を学習者に提示することが重要である。訂正の根拠情報を提示することができれば、その根拠情報に基づいて、学習者が自らの英作文に使用する単語の最終的な決定をすることを支援するとともにその語の正しい使用方法の理解に役立てることができる。

本論文では、学習者の作文支援を目的として、実際に誤り訂正システムを学習者が使用する場合を想定し、誤り訂正根拠提示の機能を備えたシステムを構築した。本研究において英作文支援の対象とする学習者は、英作文において冠詞以外

表 1: Cambridge Learner Corpus における誤りの割合 (スペル誤りを除く)

Rank	Error Type	Proportion
1	Content word choice error	19.9%
2	Preposition error	13.4%
3	Determiner error	11.7%
4	Comma error	9.3%
5	Inflectional morphology	7.4%
6	Wrong verb tense	6.7%
7	Derivational morphology	4.9%
8	Pronoun	4.2%
9	Agreement error	4.0%
10	Run-on Sentence	4.0%
11	Idiomatic Collocation and word order	3.9%
12	Confused words	1.9%
13	Conjunction error	1.7%
14	Words split with a space or joined	1.4%
15	Apostrophe error	1.3%
16	Hyphenation error	1.3%
17	Sentence fragment or two sentences that are joined	0.8%
18	Quantifier error	0.7%
19	Other punctuation error	0.4%
20	Negation formation	0.1%

の部分については十分に正しい英文を書くことのできる英語を母語としない英語学習者とする．冠詞誤り訂正の手法については既存研究を踏襲し，名詞句に対する冠詞の分類問題として冠詞誤り訂正を定式化する．分類器は，与えられた名詞句に対して周辺文脈の情報を利用してその名詞句に適した冠詞（定冠詞，不定冠詞，無冠詞のいずれか）を選択する．本研究では，それに加えて，学習者が最終的に冠詞を決定するのに役立つと考えられるいくつかの訂正根拠を提示する．

評価実験では，二つの項目について評価を行った．一つは，分類器の性能評価である．本研究では，後述の通り，冠詞選択の根拠情報として，分類に用いられた素性を提示する．従って，分類性能には寄与するが人間に理解しづらい素性を使うことができない．そのため，分類性能をできるだけ維持したまま，理解しづらい素性を除いた．評価の結果，理解しづらい素性を除いても，十分な分類性能を示すことが分かった．もう一つは，提示した訂正根拠が学習者の英作文にどれだけ貢献しているかを評価する実験である．この実験では単純な冠詞選択正解率の他に，冠詞選択に対する自信という評価項目を設けることで，訂正根拠の提示が納得して冠詞を選ぶことに貢献しているかを測定した．実験の結果，提案した訂正根拠情報を提示したときに被験者の冠詞選択正解率が向上すること，根拠情報提示前に既に正しい冠詞を選択していてもより大きな自信を伴って冠詞を選択できるようになることを確認した．

2 関連研究

2.1 英作文における誤り

英作文における誤りには，文法上誤っているもの，文法上の使用位置は正しいが語の使用方法が誤っているもの等がある．広義には，スペル誤りも含まれることがある．文法は，言語の統語構造や語形の変化を定めたルールである．例えば，前置詞句は自動詞の後に続けることができるが，名詞句は自動詞の後に続けることはできない．しかし，どの語を選択すればよいのかという使用方法は，文法では通常定義されない．例えば，冠詞の使用位置は正しくても，定冠詞，不定冠詞，あるいは無冠詞のどの冠詞を選択すればよいかは文法では定義されない．英語学習者の誤りとしては特に使用方法に関する誤りが多く，英作文の誤り訂正ではこれらの誤りを含めて訂正の課題となっている．

2.1.1 冠詞誤り

英語学習者の英作文における冠詞の誤りは学習者の誤りのなかでも最も多い誤りのひとつであり，その語彙が限定されているため計算機で扱うのに理想的な誤り訂正対象となっている．多くの冠詞誤り訂正に関する既存研究では，定冠詞，不定冠詞，無冠詞の三つの語に限定した使用方法に着目して研究が行われている．

2.2 英文の自動誤り訂正

学習者の英文に対する自動誤り訂正の研究は現在までさまざまな研究がなされている．近年では，文法誤りの自動検出・訂正システムの性能を競う評価型ワークショップ Helping Our Own (HOO) [4] が開催されるなど，高く注目されている研究分野である．このワークショップでは，実際の英語学習者の作文における誤りに対する自動誤り検出・訂正システムの性能を競う．HOO 2011 Shared Task では動詞や代名詞など多数の種類の誤りを対象としていたが，HOO 2012 Shared Task では，評価に用いられるデータの違いや対象とする誤りの範囲の違いによっ

て既存研究の間で単純に比較のできなかった既存手法の性能を比較することを目的として、対象の誤りを冠詞誤りと前置詞誤りに絞って開催されている。

冠詞誤りと前置詞誤りについてエンドユーザー向けの誤り訂正システムの実用例として、ESL Assistant [5] や Criterion Online Writing Evaluation Service [6] がある。

Criterion Online Writing Evaluation Service は冠詞誤り検出・訂正と前置詞誤り検出・訂正の両方で最大エントロピー分類器を用いる。分類器は正しく構成された英文を品詞タグ付けしたものからモデルを学習し構築される。品詞タグは冠詞や前置詞の存在する場所や存在する可能性のある場所を認識するのに使われる。

ESL Assistant は、冠詞誤り検出・訂正と前置詞誤り検出・訂正の各々で二つの最大エントロピー分類器を用いる。一つ目の分類器は与えられた文脈に対して冠詞あるいは前置詞が存在するかを判断するための分類器、もう一つは与えられた文脈にふさわしい冠詞あるいは前置詞を選択するための分類器である。分類器の学習には Criterion Online Writing Evaluation Service と同様に学習データから抽出した素性を用いる。文脈をとらえるために素性は前後 6 単語までの品詞タグと単語をもとに構成されている。さらに、ESL Assistant はユーザーの最終的な語選択の判断材料としてウェブ検索による用例を提示する機能を備えている。提示される用例はユーザーがもともと選択していた語を使用した場合とシステムにより訂正された語を使用した場合の 2 種類である。この機能は学習者が語の選択に迷ったときにしばしば行われるウェブ検索による比較を模倣したものである。用例を提示するという点で ESL Assistant は本研究の思想と類似しているが、後述する本研究の例文の検索方法はより訂正理由に則した方法となっている。

2.2.1 誤り訂正手法に関する既存研究

冠詞誤りと前置詞誤りは学習者の英文の中で主な誤りであり、冠詞と前置詞は語彙が限定されており計算機で扱いやすいため、誤り訂正に関する既存研究では冠詞誤りと前置詞誤りに焦点を当てて研究されてきた。既存研究の手法としては Memory-Based Learning [7]、決定木学習 [8, 9]、決定リスト [10]、最大エントロピーモデリング [11, 12, 13, 14, 15, 16, 5]、条件付確率場 [17]、Alternating

Structure Optimization (ASO) [18], 統計的機械翻訳 [19, 20] など, 多くの手法は教師有り学習を用いた手法である。さらにいくつかの手法では, 言語モデルの使用 [8, 21], 英語学習者による英文から構築したコーパスの使用 [11, 16], 英語母語話者による英文と英語学習者の英文との間のドメイン適応 [22, 18] などを活用した試みが行われてきた。しかしながら, これらの既存研究では訂正情報の提示のみに留まっており, その訂正の根拠を提示する手法にまでは及んでいない。

2.2.2 訂正根拠提示に関連する既存研究

一部の既存研究では訂正の根拠を提示する試みが行われている。Yi ら [21] はウェブをコーパスとして使用し, ウェブ検索エンジンによる頻度のカウントを利用することでコロケーションが正しく使われているかの判断を可能にし, 正しいようなコロケーションを実在するウェブページから示す機能を備えた英文校正システムを提案した。また, Gamon ら [8] は訂正候補に付加情報としてウェブ上に実在する用例を見せる手法を説明した。この手法は既に 2.2 節 で述べた ESL Assistant で使用されている手法である。これらの研究は用例という訂正の根拠のひとつと考えられる情報を提供しているが, 用例が学習者の英作文能力に及ぼす効果を検証できるような評価実験を行っていない。

本論文では, 訂正理由により基づいた用例を含めた訂正の根拠情報を提案し, それらが学習者の英作文能力に及ぼす効果を検証した。

3 本研究で用いる冠詞誤り訂正手法及び訂正根拠提示手法

3.1 課題設計

この研究では自動誤り訂正の根拠として冠詞の使用方法についての知識や例文を提示することを検討する．例として以下の文を考える．

A remainder of this section describes ...

この例において，下線で示した名詞句 *a remainder* の冠詞は誤りであるから，自動誤り訂正では 不定冠詞 *a* の代わりに 定冠詞 *the* が訂正候補として提示されることが期待される．この箇所の冠詞が *the* になる理由としては，*of this section* という限定が付くことにより「何の残りなのか」が明確であること，*remainder* という名詞は実体・概念との（残余という）関係を記述するものであるから，そもそも指し示すものが限定される傾向にあることが挙げられる．自動冠詞誤り訂正では，これらの理由を素性として表現することで高い信頼性をもって *the* を予測できる．さらに，分類に貢献した素性を提示することで冠詞の使用方法を説明する良い手がかりとなり，学習者が自動誤り訂正により提示された訂正候補の信頼性を検証できるようにするとともに，学習者の今後の冠詞誤りを防止する効果が期待できる．

a remainder の冠詞が *the* になりやすいことは，次のような例文からも判断できる．

- *The remainder of this paper is organized ...*
- *Formulas for the remainder term of Taylor ...*

このような例文を提示することでも学習者の自信や理解を向上させることが期待できる．

本論文では英作文支援を目的とした訂正根拠提示を備えた誤り訂正システムを構築する．訂正根拠の提示をする最初の試みとして，計算機で比較的扱いやすいタスクとして，冠詞誤りの訂正を対象にする．構築したシステムでは，冠詞の誤

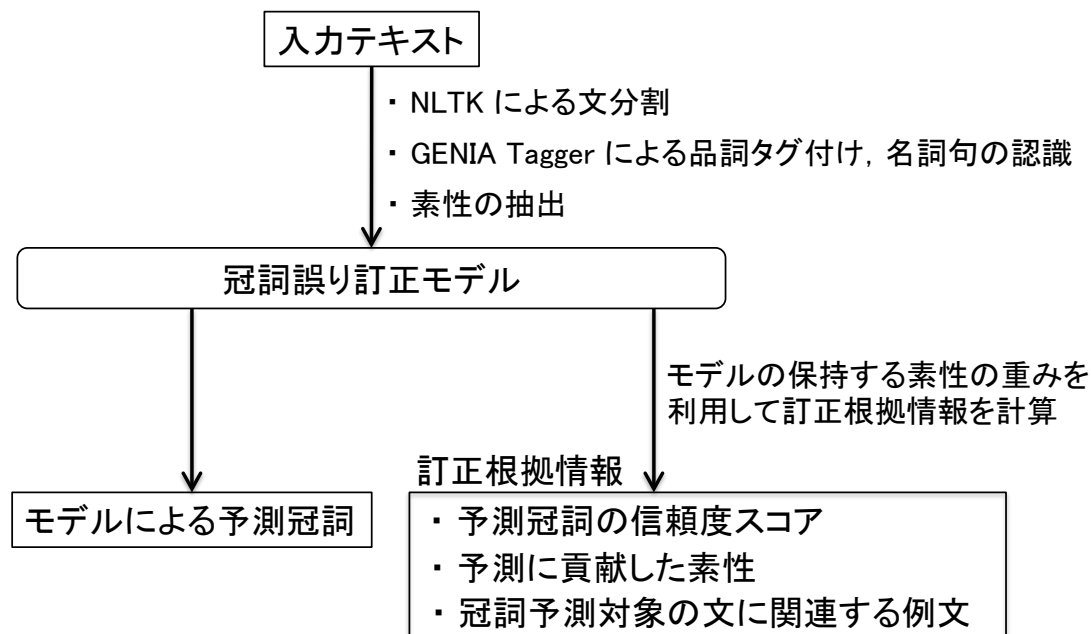


図 1: 訂正根拠提示システムの概要

り訂正候補を提示することに加えて，冠詞誤り訂正の根拠となる情報も提示する．具体的には，「システムが予測した冠詞の信頼度」，「冠詞の予測に貢献した素性」，「冠詞予測対象の文に関連する例文」の三つの種類の根拠を提示する．システムの概要を図 1 に示す．本システムでは入力テキストを前処理し，冠詞誤り訂正モデルの保持する素性の重みを利用することで訂正根拠情報の提示を実現する．本章の後続する節では，まず冠詞誤り訂正モデルについて説明した後，三つの根拠情報についてそれぞれ説明をする．

3.2 冠詞誤り訂正モデル

冠詞誤り訂正モデルの構築には，最大エントロピー法に基づく確率モデルを使用する．対象名詞句の周辺文脈 x から抽出された素性ベクトルを x ，名詞句に対応する冠詞を $y \in \{a, the, \phi\}$ とする (ϕ は無冠詞を示す)．最大エントロピー法を用いると，名詞句の文脈 x が与えられた場合の冠詞 y の条件付き確率は次式で

定義される．

$$P(y|x) = \frac{\exp(w_y \cdot x)}{\sum_{y' \in \{a, the, \phi\}} \exp(w_{y'} \cdot x)} \quad (1)$$

ここで w_y は冠詞 y を予測する場合の素性ベクトル x に対する重みベクトルである．名詞句の文脈 x において分類器が予測する冠詞 \hat{y} は以下ようになる．

$$\hat{y} = \underset{y}{\operatorname{argmax}} P(y|x) \quad (2)$$

重みベクトル w_y の学習には機械学習ツールキット Classias¹ を用いた．

3.3.2 節で詳しく述べるが、素性は訂正の根拠としてユーザーに提示することを想定しているため、分類性能を落とす事なく人間にとっての可読性の良さを確保したい．そのため、分類モデルに使用する素性は十分に単純かつ特徴的となるように設計した．例えば、複雑な素性の組合せを取り入れることは分類性能の向上には寄与するかもしれないが、人間にとっての可読性は低くなるので、素性の組合せは厳選する必要がある．

表 2 に本研究で用いた素性の一覧を示す．表 2 には各素性の説明のため次の文を用いて素性の例を示した．

Each test sentence is parsed by a bottom-up chart parser using initially the indexed subtrees only.

表 2 では下線部の *a bottom-up chart parser* という名詞句を対象にした場合の素性の例を示している．対象名詞句に付与されている不定冠詞 *a* は、正解のラベルとなるため素性抽出時には取り除かれる．本研究で用いた素性は Helping Our Own 2012 shared task [23] においてトップの性能を達成した Dahlmeier ら [3] の素性を参考に設計されている．また、本研究で用いた単語に関する素性はすべて原形を使用している．単語や品詞を用いた素性のような誤り訂正において一般的に用いられる素性の他、本論文では SPECIALIST Lexicon² から抽出した可算性の素性を用いた．

¹<http://www.chokkan.org/software/classias/>

²<http://lexsrv3.nlm.nih.gov/LexSysGroup/Projects/lexicon/2011/web/index.html>

表 2: 本研究で使した素性一覧

Feature	Example
First word in NP	bottom-up
Second word in NP	chart
Third word in NP	parser
Word 1 before NP	by
Word 2 before NP	parse
Word+POS 1 before NP	by+IN
Word+POS 2 before NP	parse+VBN
Word+POS 3 before NP	be+VBZ
Word after NP	use
Word+POS 1 after NP	use+VBG
Word+POS 2 after NP	initially+RB
Bag of words in NP	{bottom-up, chart, parser}
N-grams around article position(N=2,3,4,5)	{by_X, X_bottom-up, parse_by_X, ...}
Word before + NP	by+bottom-up_chart_parser
NP + N-gram after NP(N=1,2,3)	{bottom-up_chart_parser+use, ...}
Noun compound	chart_parser
Adjective + Noun compound	bottom-up+chart_parser
Adjective POS + Noun compound	JJ+chart_parser
NP POS + noun compound	JJ_NN_NN+chart_parser
First POS in NP	JJ
Second POS in NP	NN
Third POS in NP	NN
POS 1 before NP	IN
POS 2 before NP	VBN
POS 3 before NP	VBZ
POS after NP	VBG
Bag of POS in NP	{JJ, NN, NN}
POS N-grams around article position(N=2,3,4)	{IN_X, X_JJ, VBN_IN_X, ...}
Head of NP	parser
Countability of Head	countable
NP POS + Head	JJ_NN_NN+parser
Word before + Head	by+parser
Head + N-gram after NP(N=1,2,3)	{parser+use, parser+use_initially, ...}
Adjective + Head	bottom-up+parser
Adjective POS + Head	JJ+parser
Word before + Adjective + Head	by+bottom-up+parser
Word before + Adjective POS + Head	by+JJ+parser
Word before + NP POS + Head	by+JJ_NN_NN+parser

SPECIALIST Lexicon は生物医学分野の用語と一般的な英単語の両方の語彙情報を収録した辞書である．辞書の各語彙項目には統語論的信息や形態論的信息などが記載されている．対象名詞の語彙項目から可算性に関する情報を抽出することで素性として使用した．具体的には，対象名詞の語彙項目に記載されている可算性を表す `variants` 属性から判断をした．例えば，*student* という名詞の項目は次のように記されている．

```
{base=student
entry=E0058263
    cat=noun
    variants=reg
    compl=pphr(of,np)
}
```

この例では `variants` 属性は英語における一般的な複数形の形態（語尾に `s` を付与）をとることを示す `reg` となっているため，*student* は可算名詞であると判断する．名詞によっては，可算名詞として使用される場合と不可算名詞として使用される場合の両方の可能性を持つのでそのような名詞については辞書に記載されている情報からは可算性が判断できないため，可算性に関する素性は用いていない．そのような例として *method* という名詞の語彙項目を次に示す．

```
{base=method
entry=E0039936
    cat=noun
    variants=reg
    variants=uncount
}
```

この例では，可算名詞であることを示す `reg` と不可算名詞であることを示す `uncount` の両方の `variants` 属性を含むため，*method* という名詞については可算性に関する素性は用いないということになる．

また，実行時間を減らす目的と過学習を軽減する目的で，トレーニングデータ中に 1 度しか出現しない素性については取り除いて学習を行った．

誤り訂正モデルは冠詞誤りとそれらの訂正情報が含まれた学習者コーパス (e.g., Cambridge Learner Corpus (CLC)³, NICE⁴, The NICT JLE Corpus⁵) から直接学習することが望ましいが，そのような学習者コーパスはサイズが小さく，機械学習に用いるには十分な学習事例が確保できないという問題がある．そのため，学習者コーパスを用いる代わりに，大規模かつ正しく書かれた英文から学習したモデルを構築する．

冠詞誤り訂正モデルを学習するときは，正しい英語が書かれている文書を Natural Language Toolkit (NLTK) [24] を用いて文単位に分割する．さらに GENIA tagger [25] を用いて，品詞のタグ付けとチャンキングを行うことにより，名詞句を認識する．認識されたそれぞれの名詞句に対して，名詞句の冠詞を取り除くことで学習事例を作成する⁶．もともと名詞句に付与されていた冠詞に基づいて，各学習事例には正解ラベル $y \in \{a, the, \phi\}$ が割り当てられる．

表 2 で用いた例文を入力として与えたときの GENIA tagger の出力は次のようになる．

Each	Each	DT	B-NP	0
test	test	NN	I-NP	0
sentence	sentence	NN	I-NP	0
is	be	VBZ	B-VP	0
parsed	parse	VRB	I-VP	0
by	by	IN	B-PP	0
a	a	DT	B-NP	0
bottom-up	bottom-up	JJ	I-NP	0

³<http://www.cambridge.org/gb/elt/catalogue/subject/custom/item3646603/Cambridge-International-Corpus-Cambridge-Learner-Corpus/>

⁴<http://sugiura5.gsid.nagoya-u.ac.jp/~sakaue/nice/>

⁵http://alaginrc.nict.go.jp/nict_jle/

⁶品詞タグとして PRP, PRP\$, WDT, WP, WP\$ のいずれかを含む名詞句は基本的に冠詞を伴わないため学習事例の対象から外す．同様に，冠詞以外の限定詞 (e.g., *this*, *that*) を含む名詞句も学習事例の対象から外す．

chart	chart	NN	I-NP	0
parser	parser	NN	I-NP	0
using	use	VBG	B-VP	0
initially	initially	RB	B-ADVP	0
the	the	DT	B-NP	0
indexed	indexed	JJ	I-NP	0
subtrees	subtree	NNS	I-NP	0
only	only	RB	B-ADVP	0
.	.	.	0	0

入力文は単語分割され，各単語の情報は，表層形，原形，品詞タグ，チャンクタグ，固有表現タグの順にタブ区切りで出力される．チャンクタグは IOB2 フォーマットで表現され，I はそのチャンクの先頭以外の単語，O はそのチャンクに含まれない単語，B はそのチャンクの先頭の単語であることをそれぞれ意味する．チャンクタグとして B-NP が付与された単語とそれに続く I-NP が付与された単語のまとまりを名詞句として認識する．つまり，この例では，*Each test sentence* と *a bottom-up chart parser* と *the indexed subtrees* が名詞句として認識される．

a bottom-up chart parser を例にいくつかの素性を説明する．まず，名詞句に付与されている冠詞は正解のラベルとなるため取り除かれる．つまり，ここで素性抽出の対象となる名詞句は *bottom-up chart parser* である．この名詞句で「First word in NP」という素性を考えると，名詞句の最初の単語である *bottom-up* が GENIA tagger の出力の 2 列目の原形から抽出される．同様に，「First POS in NP」は GENIA tagger の出力の 3 列目の品詞タグから抽出され，JJ (形容詞) となる．「Noun compound」は複合名詞を抽出する素性で，名詞句内の名詞接続を NN (一般名詞単数形) と NNS (一般名詞複数形) の品詞タグをもとに抽出する．組合せの素性としては，例えば，名詞句内の品詞列を表す素性「NP POS」と名詞句内の複合名詞を表す「Noun compound」を組み合わせた「NP POS + Noun compound」がある．*bottom-up chart parser* の例では，この素性の内容は *JJ_NN_NN+chart_parser* となる．このような簡潔な組合せ素性であれば，訂正性能の向上に貢献しつつ，人間にとっての可読性も保つことができる．

3.3 冠詞誤り訂正根拠情報

既に述べたように、本研究では、冠詞の誤り訂正候補を提示することに加えて、冠詞誤り訂正の根拠となる情報として、「システムが予測した冠詞の信頼度」、「冠詞の予測に貢献した素性」、「冠詞予測対象の文に関連する例文」の三種の根拠を提示する。

「システムが予測した冠詞の信頼度」は既存研究ではシステムが内部に保持していた予測の信頼度を学習者に見せることで判断の指標にする。

「冠詞の予測に貢献した素性」は訂正モデルによる冠詞予測の理由を説明するものとして提示する。本研究における訂正モデルの素性は名詞句の周辺文脈の情報を表したものであるから、冠詞予測に貢献した素性は冠詞決定との関連が特に強いと考えられる。人手で全て書き尽くすことが非現実的である冠詞を決定づけるルールをこの素性により近似的に表現する。

「冠詞予測対象の文に関連する例文」は、2.2 節で述べた既存のエンドユーザー向けの誤り訂正システム *ESL Assistant* でも用いられているような例文を冠詞選択の判断の基準の一つとする。本研究において用いる例文は単純な文字列検索を行って取得するのではなく、誤り訂正モデルが保持する重みを利用することによって、訂正根拠としてより有用であると考えられる例文を取得する。

3.3.1 訂正モデルによる冠詞予測の信頼度の提示

本論文では分類モデルにより予測された冠詞 \hat{y} に加えて、訂正根拠情報のひとつとして条件付き確率 $P(y|x)$ を提示する。この根拠情報は訂正モデルによって予測された冠詞の信頼度としてユーザーに提供される。モデルの予測した冠詞の信頼度が高ければ、ユーザーはモデルによる予測を信頼することができ、モデルの予測した冠詞の信頼度が低ければユーザーは自身の冠詞選択を尊重することができる。

例として後述する評価実験 (4 章) で構築したモデルを使用した場合の次の例に対する信頼度スコアを考える。

We also proposed *method* to further integrate the synonym relations to the LM approaches.

この例の下線部 *method* に付与される可能性のある各冠詞に対する信頼度スコアは不定冠詞 *a* が 0.9934 , 定冠詞 *the* が 0.0017 , 無冠詞が 0.0049 である . このときモデルによる予測冠詞は不定冠詞 *a* となるが , 信頼度スコアから予測に高い信頼性があると判断することができる . もしこのとき , 予測冠詞である不定冠詞 *a* の信頼度スコアが他の冠詞の信頼度スコアと拮抗しているのあれば , モデルの予測が誤っている可能性も高くなっていると判断できる .

3.3.2 訂正モデルによる冠詞予測の理由の提示

なぜ訂正モデルが与えられた名詞句 x に対して冠詞 \hat{y} を選択するのかという理由を示したい . これをするために , 本研究では分類に最も貢献したと考えられる素性を示す . 冠詞の誤り訂正モデルは名詞句に付与される各冠詞それぞれの場合について素性 x_i に対する重みとして $w_{a,i}$, $w_{the,i}$, $w_{\phi,i}$ の値を保持している . そこで素性 x_i が冠詞 y を推薦する強さ $d(i, y)$ を次式で求める .

$$d(i, y) = \frac{\exp(w_{y,i})}{\exp(w_{a,i}) + \exp(w_{the,i}) + \exp(w_{\phi,i})} \quad (3)$$

本研究では , それぞれの冠詞 y に対して , 獲得した $d(i, y)$ の値が高い 3 つの素性を選択し , 訂正根拠情報のひとつとした . 実際にユーザーに根拠情報として提示するときには , 素性の表現は英語学習者にとって変換ルールによってわかりやすい表現に言い換えて表示をした . 訂正モデルにおいて , 素性は素性のタイプ (e.g., HEAD-OF-NP) とその内容 (e.g., *method*) で構成されている . そのため , 素性のタイプ毎に素性の内容部分を当てはめるスロットを設けたテンプレート表現を用意し , 内容に応じてスロットを埋めることで人が読みやすい表現に言い換えるルールを用意した . 例えば , 素性タイプ HEAD-OF-NP についての変換ルールは「“HEAD-OF-NP= arg ” \rightarrow “The head of the NP is arg ”」となっている (arg は素性内容を入れるスロット) . このとき , “HEAD-OF-NP= $method$ ” という素性はテンプレートのスロットを素性の内容である *method* で埋めることで “The head of the NP is *method*” という表現に変換される .

3.3.1 節と同様に以下の文を具体例として考える．

We also proposed method to further integrate the synonym relations to the LM approaches.

下線部の名詞句 *method* に付与されるべき冠詞は不定冠詞 *a* であり，モデルによる予測冠詞も *a* となるが，不定冠詞 *a* の場合の $d(i, y)$ の値が高い三つの素性は「*propose* *X* という単語列を含む (*X* は冠詞を表す)」，「名詞句の 1 語前の単語が *propose*」，「名詞句の前の単語が *propose*，かつ，名詞句の主辞が *method*」である．これらの素性から学習者は「動詞 *propose* の目的語となる名詞句は不定冠詞をとりやすいこと」や「動詞 *propose* の目的語が名詞 *method* となる場合に不定冠詞をとりやすいこと」が訂正の理由として読み取れる．これらの訂正理由は冠詞選択の判断になるとともに，今後同様の誤りを防ぐ冠詞の使用方法の知識にもなると考えられる．

3.3.3 冠詞予測理由に基づく例文の提示

本研究では，対象名詞句の周辺文脈に基づいた例文も訂正根拠情報として提示する．例文は以下の手順で検索した．

この手順は対象名詞句に付与される各冠詞それぞれの場合について実行される．

1. システムは 3.3.2 節と同様の方法で，対象名詞句における素性から $d(i, y)$ の値が高い素性を三つの素性を例文検索のキーとなる素性として選択する．システムはキーとなる三つの素性を持ち，かつ，対象名詞句と同じ主辞を持つ名詞句が含まれる文を例文候補として学習データから抽出する．
2. システムは対象名詞句の素性ベクトルと各例文候補の素性ベクトルのコサイン類似度を計算する．コサイン類似度の値が高い三つの文を最終的に提示する例文とする．

3. もしシステムが三つの例文を見つけることができなかった場合，システムはキーと鳴る素性の数を一つずつ減らしながら同様の手順を三つの例文が獲得できるまで繰り返す．

上記例文検索手順を 3.3.1 節と同様に以下の文を例として考える．

We also proposed method to further integrate the synonym relations to the LM approaches.

この例において，後述する評価実験 (4 章) で使用するモデルによる予測冠詞である不定冠詞 *a* に関する上記例文検索手順によって得られる例文は次の三つである⁷．

- *We also proposed a method to find paraphrases in the context of two Named Entity instances in a large un-annotated corpus.*
- *In her paper, she also proposed a method to automatically distinguish between cognates and false friends, while examining the performance of seven different machine learning classifiers.*
- *One of the authors previously also proposed a method to mine discussion records using XML annotations and a method to find important remarks in a discussion thread based on the reply-to structure and participants' opinions.*

このような例文を提示することで，学習者の冠詞選択の判断材料を増やすとともに実例を見ることでより実際的な冠詞の使用方法を身につけることを期待する．

⁷4 章で詳しく述べるが，使用したデータの都合上これらの例文は ACL Anthology に収録されている論文からの引用となる．

4 評価実験

提案手法を評価するために，二つの評価実験を行った．ひとつは冠詞誤り訂正モデルの分類正解率を実証する実験である．もうひとつの実験は提案手法による冠詞予測と訂正根拠情報を提示した際の英語学習者の冠詞選択能力を評価する実験である．

4.1 実験に使用したデータセット

テクニカルライティングを想定して，実験には自然言語処理分野の論文で構成されたデータを用いた．テクニカルライティングを想定する理由は，論文として投稿されているものはよく推敲されており，誤りのない正しい英文であると考えられるため，また，冠詞が一意に定まることが多いためである．自然言語処理分野の論文で構成されたコーパスとしては ACL Anthology Reference Corpus [26] が存在するが，このコーパスには OCR による文字の誤認識が多く含まれている．そこで，本研究では，分類器の性能評価実験と被験者による評価実験の両方で，ACL Anthology⁸ のウェブサイト⁸ に収録されている 2000 年から 2012 年の論文データをクロールすることで獲得したデータセットを用いた．ダウンロードした PDF ファイルからプレーンテキストを抽出して，著者名や節の見出しなど不要な部分を取り除いた．両方の実験において，テストデータとトレーニングデータに重なりのないように，ACL 2012 で発表された論文データから抽出された部分をテストデータとして用いて，それ以外のデータをトレーニングに使用した．トレーニングデータは 3,443,940 文で冠詞誤り訂正のトレーニング事例の数は 9,468,343 個となった．テストデータには 73,430 文が含まれ，テスト事例となる名詞句の数は 168,900 個であった．

⁸<http://aclweb.org/anthology-new/>

4.2 実験に使用した分類器の性能の評価

名詞句に対する冠詞の分類性能について評価実験を行った．全ての事例を無冠詞 ϕ に分類するベースラインの正解率は 65.02% である．分類器の学習にはトレーニングデータの全てを利用し，テストデータに対する冠詞分類の正解率を計測した．3 種類の冠詞の分類の正解率は 86.35% で，ベースラインを大きく上回る性能を示した．表 3 に本評価実験と同様に正しい英文に対して冠詞を予測する問題を解いている既存研究の性能の一覧を示す．評価に使用しているデータや評価方法の違いがあるため厳密な比較はできないが，構築したモデルの冠詞予測性能は，既存研究と比較しても引けを取らないことが確認できる．限られた素性に絞って学習した誤り訂正モデルであるにもかかわらず，高い分類性能を示したことから，冠詞予測に重要な情報をとらえた素性の選択ができているといえる．また，評価の対象となる英文を自然減言語処理分野の論文と限定したことにより，複雑な言い回しが少なくなったことも，分類性能の向上に寄与していると考えられる．

表 3: 既存研究の正しい英文上での冠詞予測性能

Citation	Performance
Minnen <i>et al.</i> [7]	84% accuracy
Lee [12]	88% accuracy
Han <i>et al.</i> [6]	88% accuracy
Lee and Seneff [27]	86% precision, 76% recall
Nagata <i>et al.</i> [28]	83% precision, 65% recall (on article errors in prepositional phrases)
Turner and Charniak [29]	87% accuracy
De Felice and Pulman [14]	92% accuracy
Gamon <i>et al.</i> [8]	86% accuracy

4.3 被験者による評価

提案手法の有効性を検証するために提案手法によって提示される根拠情報と訂正情報を被験者に与えたときの被験者の冠詞選択能力を評価する実験を行った。この実験では、被験者には 1 文が提示され、被験者はその文中の指定された名詞句について適した冠詞を選択する。また、提案手法のいくつかの根拠情報を提示した場合での被験者の冠詞選択能力を計測した。実験を行った被験者は英語を第二言語とする日本語を母語とする学生 4 名と中国語を母語とする学生 1 名で、いずれの被験者も本システムの開発には携わっていない。

実験手順を以下に示す。

1. 被験者には、冠詞が取り除かれた名詞句が含まれた 1 文が与えられる。被験者はまず文のみを閲覧して、その名詞句にふさわしいと思われる冠詞を選択する。
2. システムにより冠詞誤り訂正モデルの各冠詞に対する信頼度スコアが付加情報として提示される。このとき被験者は自身の冠詞選択を (1) の時点で選択した冠詞から変更することが許される。
3. さらに加えて、訂正の理由となる素性と例文の根拠情報が被験者に提示される。被験者には (2) の時点で選択した冠詞を変更する機会が与えられる。
4. 1-3 のサイクルを全てのテスト事例に対して被験者の回答が揃うまで繰り返す。

上記手順 1-3 では、選択された冠詞の他、被験者は選択した冠詞に対する自信を記録する。実験の途中で、被験者が正解である冠詞を閲覧することはできない。

文のみを閲覧したときの正答率と比較して、根拠情報を閲覧したときに、どの程度正解率が向上するかを調べた。実験のなかで、根拠情報のひとつである各冠詞の予測確率のみを閲覧する段階を設けたのは、その情報が学習者の冠詞予測正解率に最も貢献すると考えられるからである。根拠情報を閲覧したときの正解率の向上が大きければ、訂正の根拠情報を見せるという提案手法の有効性を確認で

きる．また，被験者の冠詞予測に対する自信が適切に向上していれば，提案手法の有効性を確認できる．

実験に使用するデータは 4.1 節で説明したテストデータのうち，248 個の事例を使用した．これらの事例は，テストデータから無作為に抽出したインスタンスから以下の条件に当てはまる事例を手で取り除いたものである．

- 対象名詞句が記号で囲まれている
- 対象名詞句に数字，固有名詞，数式，“there” が含まれる
- 名詞句が誤って抽出されている
- 対象名詞句あるいは周辺文脈にスペル誤りなどの誤りが含まれる
- 対象名詞句の含まれる文が短すぎる

これらの条件は被験者とシステムともに予測の妨げとなる事例を取り除く目的で構成されている．

この評価に使用したデータにおける分類器の正解率は 85.48% であった．分類器の信頼度スコアの 30 事例毎の平均に対しての分類器正解率の移動平均正解率を図 2 に示す．図 2 から分類器の信頼度スコアが実際の正解率に概ね比例しており，信頼度スコアが被験者の冠詞選択の妨げにならないことが確認できる．また，信頼度スコアが高いほど，他の根拠情報の質も向上すると考えられる．

図 3 は構築したシステムの出力例である．図の左側最初の縦列には分類器の信頼度スコアのパーセンテージが表示されている．そのすぐ右の列には対応する冠詞が表示されている．3 番目の列の上部には冠詞予測の対象となる文が示されている．残りの行では，根拠情報である訂正の理由となる素性と例文が示されている．実験を行うとき，各々のテスト事例の初期状態では対象となる文と冠詞の部分だけが表示されている．被験者は対象文のみを閲覧し冠詞を選択した後，図上部の「Confidence」と記されたボタンをクリックすることで分類器の信頼度を閲覧することができる．付加情報として信頼度を閲覧し冠詞を選択した後，同様に，「All」ボタンをクリックすることで 3 番目の列に残りの根拠情報が表示されるので，それらの情報を閲覧し再度冠詞を選択する．

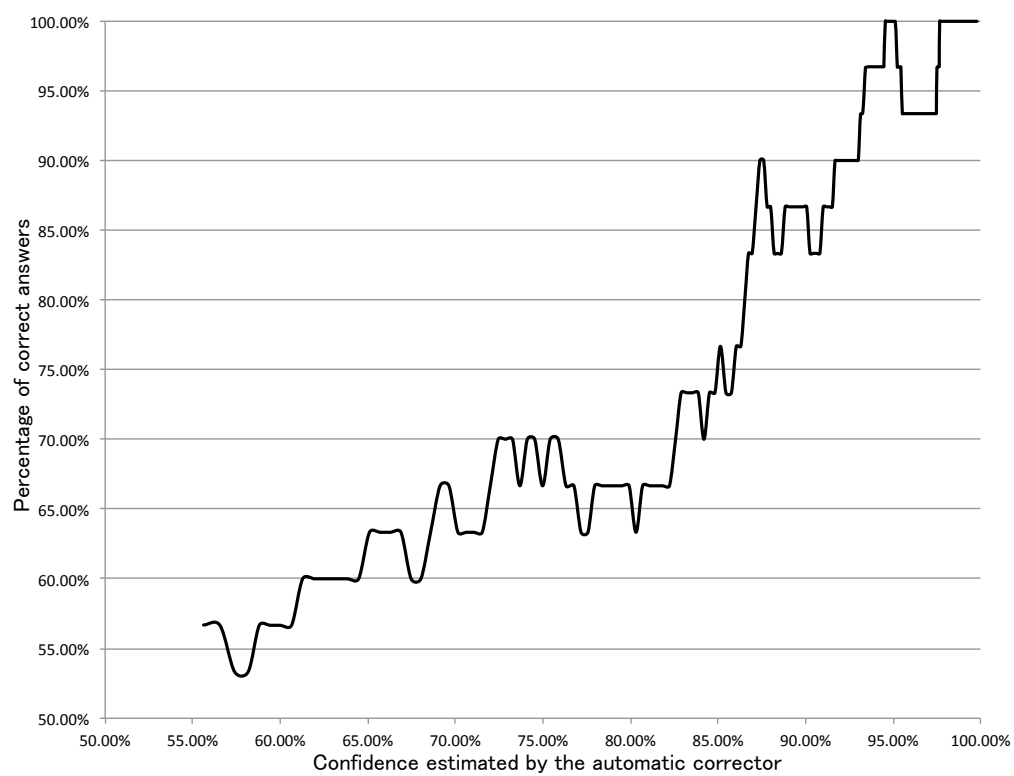


図 2: 分類器の信頼度スコアと正解率との関連

Confidence		All
		We also proposed method to further integrate the synonym relations to the LM approaches .
99.34%	a/an	<ul style="list-style-type: none"> Because "propose_[ARTICLE]" is included in the sentence. Because the word before NP is "propose". Because the word before NP is "propose" and head of NP is "method" <ul style="list-style-type: none"> We also proposed a method to find paraphrases in the context of two Named Entity instances in a large un-annotated corpus (Sekine 05) . In her paper , she also proposed a method to automatically distinguish between cognates and false friends , while examining the performance of seven different machine learning classifiers . One of the authors previously also proposed a method to mine discussion records using XML annotations (Murakami et al. , 2001) and a method to find important remarks in a discussion thread based on the reply-to structure and participants' opinions (Murakami et al. , 2007) .
0.17%	the	<ul style="list-style-type: none"> Because whole POS in NP is "noun(singular or mass)" and head of NP is "method". Because head/POS of NP is "method/noun(singular or mass)". Because "verb(past tense)_[ARTICLE]_noun(singular or mass)" is included in the sentence. <ul style="list-style-type: none"> We also applied the method to 200 compound nouns of technical terms (Aiso , 1993) in English . I also extended the method to automatically identify narrative focalizers , and showed that it is possible to do so with near-perfect accuracy . McIlroy used the method to compress a dictionary of 32 , 000 words into a PDP-11 address space of 64k bytes .
0.49%	Ø	<ul style="list-style-type: none"> Because POS 2 before NP is "adverb". Because "adverb_verb(past tense)_[ARTICLE]_noun(singular or mass)" is included in the sentence. Because head of NP is "method" and "to_further" continue after NP. <ul style="list-style-type: none"> We also found regression methods to be the most efficient of the statistical models with which we experimented . In this paper we unify the probability models for phrasal translation with the algorithms for synchronous parsing , harnessing the benefits of both to create a statistically and algorithmically wellfounded method for phrasal analysis of bitext . We also tested anmeasure other method that exploited the confidence of classification but the results were not satisfactory .

図 3: 構築したシステムの出力例

表 4: 各被験者の冠詞予測正解率

	文のみ提示	+信頼度スコア提示	+全根拠情報提示
被験者 1	59.27%	85.89%	85.48%
被験者 2	64.92%	85.89%	81.05%
被験者 3	37.10%	79.03%	77.42%
被験者 4	62.10%	79.84%	83.87%
被験者 5	64.52%	85.08%	84.27%

実験の結果として各評価者の正解率を表 4 に示す．全ての評価者で分類器確率を閲覧した段階で分類器性能と同等程度の正解率であった．いくつかの被験者では分類器の性能を超える正解率であった．分類器の性能が完全でない以上，学習者が盲目的にときおり誤ることのある分類器の予測に従うことはあり得ないと考えられるため，この結果は分類器予測確率を見せるだけでも冠詞選択の支援となることを示唆している．

4.4 考察

前説で述べたように，提案手法の根拠情報の一つである分類器の信頼度スコアが学習者の冠詞予測能力を向上させることが確認できた．しかし，残りの二つの根拠情報を合わせて提示しても，被験者の予測正解率にほとんど変化は見られなかった（1 人の被験者ではわずかに上昇し，残りの被験者ではわずかに減少した）．提案手法の根拠情報は全て分類モデルの素性の重みに基づいているため，信頼度スコアを見せた時点で正解の冠詞に導かれる事例が多かったため，正解率の変化が見られなかったのではないかと考えられる．そこで，我々の根拠情報の有効性を図るため，信頼度スコアを提示する段階から全根拠情報を提示する段階へ移る部分について，以下の定義に従って各被験者結果からテスト事例を分類した．

- ポジティブな根拠情報
 - － 不正解から正解に選択が変更された事例

- － 正解から正解に選択が変更され，かつ，選択した冠詞に対する自信が向上した事例
 - － 不正解から不正解に選択が変更され，かつ，選択した冠詞に対する自信が減少した事例
- ニュートラルな根拠情報
 - － 正解の選択から変更がなく，かつ，自信の変化を伴わなかった事例
 - － 不正解の選択から変更がなく，かつ，自信の変化を伴わなかった事例
- ネガティブな根拠情報
 - － 正解から不正解に選択が変更された事例
 - － 不正解から不正解に選択が変更され，かつ，選択した冠詞に対する自信が向上した事例
 - － 正解から正解に選択が変更され，かつ，選択した冠詞に対する自信が減少した事例

この分類に則って各被験者のテスト事例についてカウントした結果を表5と図4に示す．表5と図4から，自信は被験者の主観に基づくので被験者により自信の変化のしやすさにばらつきが見られるが，全ての被験者でポジティブな根拠の数がネガティブな根拠の数を大きく上回っていることが分かる．このことから，確率のみを提示する場合よりも，訂正の理由となる素性と例文も合わせて提示することで被験者が納得して正しい冠詞を選択することに寄与することが確認できた．また，素性や例文は被験者に冠詞の用法を示すことができるので，学習者が今後同様の誤りをするのを防ぐことができるようになることや，もともと学習者自身の冠詞選択が正しかった場合であってもさらに冠詞について知識を深めることなど，訂正情報のみや信頼度スコアを見せるだけよりも学習効果を与えることができる．分類モデルの構築に使用する素性をさらに精巧に作り込むことで分類器そのものの正解率の向上と共に学習者の支援につながるより良い根拠情報を出すことができると考えられる．

表 5: 根拠情報の有効性

	ポジティブ	ニュートラル	ネガティブ
被験者 1	79.44%	6.85%	13.71%
被験者 2	48.79%	29.44%	21.77%
被験者 3	18.15%	72.58%	9.27%
被験者 4	83.47%	0.40%	16.13%
被験者 5	81.05%	4.84%	14.11%

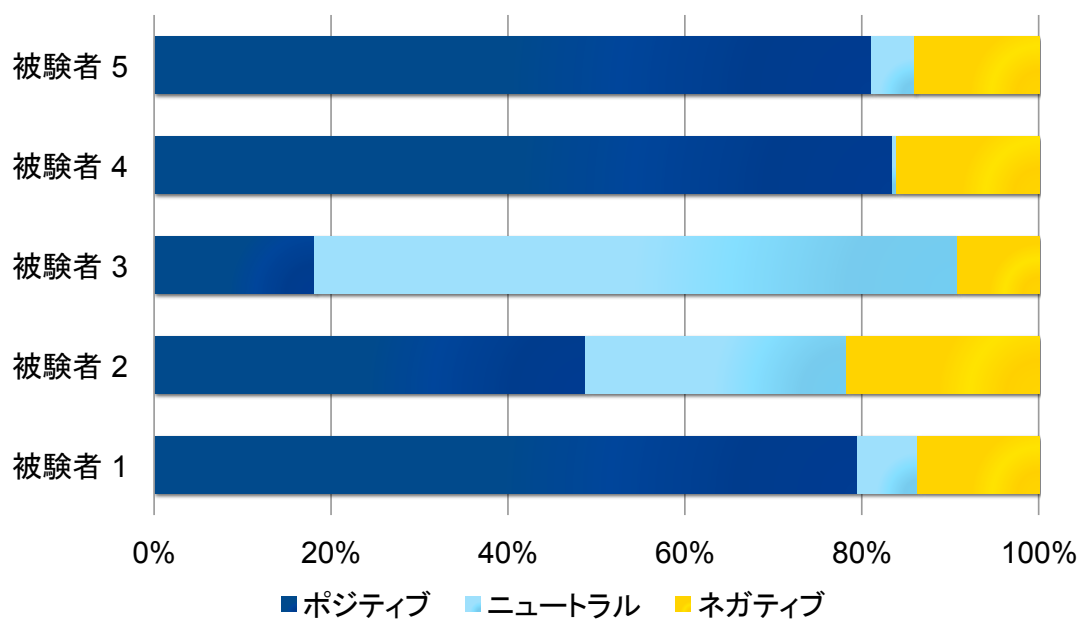


図 4: 根拠情報の有効性

次に，テスト結果を詳細に分析し，本研究の改善点を探求する．まず，訂正システムが正しいときに根拠提示によって被験者がシステムによる訂正をより信じていることができるか，システムが訂正を誤ってしまったときであっても根拠提示によってシステムの訂正に惑わされずに被験者が正しく冠詞を予測できるようになるかを調査した．表 6 は被験者が文のみを閲覧して冠詞を選択したときと根拠情報を閲覧して冠詞を選択したときの，システムによる予測冠詞が正しいか誤っているか，被験者の選択冠詞が正しいか誤っているかで分類をしたときの分布の変化である．表 6 から根拠提示をしない場合とした場合とを比較すると，システムが適切な冠詞を正しく予測できているときは根拠提示によって被験者が正しく冠詞を選ぶことができることがわかる．しかし，システムが適切な冠詞を正しく予測できていない場合は根拠提示によって被験者が誤った冠詞を選択してしまう傾向にある．これは，本研究で提案した根拠情報は全てモデルの保持する素性の重みに基づいて設計されているため，冠詞予測にうまく適応しなかった素性の重みを利用しても有効な訂正根拠が提示できないことが原因であると考えられる．

表 6: 根拠提示無しから根拠提示有りでの予測分布の変化

	システム正，被験者誤		システム正，被験者正		システム誤，被験者正		システム誤，被験者誤	
被験者 1	33.47%	2.02%	52.02%	83.47%	7.26%	2.02%	7.26%	12.50%
被験者 2	28.63%	10.48%	56.85%	75.00%	8.06%	6.05%	6.45%	8.47%
被験者 3	54.84%	11.69%	30.65%	73.79%	6.45%	3.63%	8.06%	10.89%
被験者 4	29.44%	4.84%	56.05%	80.65%	6.05%	3.23%	8.47%	11.29%
被験者 5	28.23%	5.65%	57.26%	79.84%	7.26%	4.44%	7.26%	10.08%

システムによる予測冠詞は正しかったが，最終的に被験者が冠詞選択を誤ってしまった事例について詳しく分析する．このような事例は，各被験者でそれぞれ，5 事例 (2.02%)，26 事例 (10.48%)，29 事例 (11.69%)，12 事例 (4.84%)，14 事例 (5.65%) 見られた．5 人の被験者のうち 3 人以上が最終的に誤った冠詞を選択してしまった事例は，テストに使用した 248 事例のうち 7 事例であった．これらの 7 事例を調べたところ，次の 4 つの傾向が見られた．

傾向 1 信頼度スコアが拮抗している (システムの予測冠詞と次点の冠詞の信頼度の差が 50 ポイント以内)

傾向 2 単語や文の区切りが誤っている例文が提示されている

傾向 3 1 文からでは冠詞を決定するのが困難

傾向 4 根拠提示はできているが、誤りである冠詞側にも類似した例文がある

この傾向によって各事例を分類した結果を表 7 に示す．システムの予測冠詞の信頼度スコアと 2 番目に信頼度スコアの高い冠詞の信頼度スコアを比較すると，その差が 50 ポイント以下の事例が 7 事例のうち 5 事例あり，冠詞予測の信頼度スコアが拮抗しているために，これらの 5 事例は最終的な冠詞の決定が難しい事例であったと考えられる．これらの事例のうち，二つの事例のシステム出力を図 5 と図 6 に示す．

表 7: 被験者が誤ってしまった事例の分類

	事例 1	事例 2	事例 3	事例 4	事例 5	事例 6	事例 7
傾向 1							
傾向 2							
傾向 3							
傾向 4							
その他							

図 5 は，*intermediate results* という名詞句に付与されるべき冠詞をシステムが定冠詞 *the* と予測した例である．テストデータ中に実際に付与されていた冠詞は定冠詞 *the* であるため，システムは正しい冠詞を予測できている．しかし，多くの被験者が最終的に誤った冠詞である無冠詞を選択した．対象名詞句 *intermediate results* の周辺文脈を見てみると，名詞句の前には *based on*，後には *obtained from* が続く．ここで，対象名詞句 *intermediate results* に *the* が付与される重要な根拠は，後に続く *obtained from* によって *results* の範囲が限定されるからである．システムが定冠詞 *the* の根拠として提示した根拠にはこれに関連した「名詞句の次の単語が *obtain*」という素性が提示されており，例文もそれに準じたものが提示されている．それにも関わらず，被験者が最終的に無冠詞を選択してしまったのは，対象名詞句 *intermediate results* の前に出現する *based on* という文脈が反

映された例文が提示されているので直感的に無冠詞が正しいと感じたのだと考えられる．この例では，想定通りの根拠提示ができていますが，このように被験者が誤ってしまう例を減らすためには，より良い説明を可能にする素性表現を追加するなど，学習に使用する素性を充実させる必要がある．

図 6 は，*feature* という名詞句に *the* が付与されると予測した例である．この例は，テスト事例のなかでも特に各冠詞の信頼度スコアが均一に近い例である．このように各冠詞の信頼度スコアが似ているときは被験者に冠詞選択の判断の基準となる素性や例文を提示することを想定しているが，この例では十分な根拠が出せていない．例えば，提示した例文のなかには，単語や文の区切りが適切に表現されていないものとして次の例文がある．

The feature wh ant signals the presence of a wh-phrase in the antecedent utterance, which seems to be the best cue for classifying Short Answers.

下線部分 *wh ant* は元データである論文では *wh_ant* と一つの語となっているが，文を抽出した PDF ファイル内部のテキスト情報が誤っているために二つの語に分かれてしまったものである．システムがこの文に GENIA tagger によって品詞タグ付けを行うときに，*wh* が前置詞としてタグ付けされてしまい，類似度の高い例文として出現している．このように可読性の悪い例文を提示してしまうと，システムの予測した冠詞の信用を落とす要因の一つになると考えられる．

		A meta-classifier decides the final output based on intermediate results obtained from several base classifiers .
0.19%	a/an	<ul style="list-style-type: none"> base_on_[冠詞] という単語の並びを含むため 名詞句の前の単語が on, かつ, 形容詞_普通名詞複数形 が含まれる, かつ, 名詞句末の単語が result のため 名詞句の最初の単語の品詞 が 形容詞
		<ul style="list-style-type: none"> This method introduces richer information to the chunker than previous methods that base on a single morphological analysis result . The next proposition , which is based on a result in Graca et al . In its turn , any MRG operation performed on a combo-model results again in a valid combo-model .
69.84%	the	<ul style="list-style-type: none"> [冠詞]_形容詞_普通名詞複数形 という品詞の並びを含むため 名詞句内2語目の単語 が result 名詞句の次の単語 が obtain
		<ul style="list-style-type: none"> The conclusions are based on the final results obtained after both the grammar and machine learning have been applied (table 6) . In the following paragraphs we will discuss the quality of the trained substitution probabilities as well as comment on the dialectological results obtained with the trained model . The experimental results obtained from this research demonstrate that it is possible to apply a syntacticbased approach to deduce textual entailment from a text-hypothesis pair .
29.98%	Ø	<ul style="list-style-type: none"> 名詞句末の単語の品詞 が 普通名詞複数形 名詞句の2語目の単語が base/過去分詞 であるため 名詞句内2語目の単語の品詞 が 普通名詞複数形
		<ul style="list-style-type: none"> Based on intermediate results from a CRF tagger , which employs local features and global features , we use class-based HMM to revise the substrings with low marginal probabilities . Based on quantitative results (Pk and WindowDiff metrics) use of specific algorithm was recommended and impact of pre-processing strategies was assessed . Based on prior results for monologue , we selected pitch and amplitude features for consideration .

図 5: システムの予測は正しいが , 最終的に被験者が冠詞選択を誤った事例 (1)

		Feature from the previous example is now 'PP pobj 1997' .
34.14%	a/an	<ul style="list-style-type: none"> 名詞句全体の品詞の並びが 普通名詞, かつ, 名詞句末の単語が feature のため 名詞句末の単語が可算名詞 名詞句末の単語の品詞 が 普通名詞
		<ul style="list-style-type: none"> A feature of the CSM-based method is that it can extract not only pairs of related words but also sets of related words because it connects tuples consistently . A feature of CSM is its ability to estimate hierarchical relations between words . A feature on an arc $e2 \in E2$ can access local characteristics of the confidence-rated superposition of the L1 's outputs , in addition to the information of the input x .
43.92%	the	<ul style="list-style-type: none"> 名詞句末の単語が feature/普通名詞 であるため [冠詞]_feature という単語の並びを含むため 名詞句の最初の単語 が feature
		<ul style="list-style-type: none"> The feature wh ant signals the presence of a wh-phrase in the antecedent utterance , which seems to be the best cue for classifying Short Answers . The feature of reflexive coreference is easier to approximate , as this coreference takes place within the same clause . The feature of grammatical gender was specifically isolated because it is rarely salient for cross-language applications such as machine translation (where grammatical gender rarely transfers) , and because its induction algorithm in Section 4.1 depends heavily on the size of the monolingual corpus (which is small in these experiments , suggesting size-dependent potential for significant further improvement here) .
21.94%	Ø	<ul style="list-style-type: none"> 名詞句末の単語 が feature 名詞句の次の単語の品詞 が 前置詞 名詞句全体の単語の並びが feature, かつ, 名詞句の後に from_previous が続くため
		<ul style="list-style-type: none"> e are features from a previous event . Since BANNER utilizes features from previous research on GENETAG , YAPEX and JNLPBA , we expect good performance on these data sets . The third semantic algorithm combines features from the previous two .

図 6: システムの予測は正しいが , 最終的に被験者が冠詞選択を誤った事例 (2)

5 おわりに

5.1 まとめ

本稿では、冠詞の誤り訂正という課題に対して、誤り訂正システムの訂正性能を落とすことなく、最終的に学習者が冠詞を選ぶまでのトータルの支援として、訂正の根拠として、分類器の各冠詞に対する訂正信頼度スコア、冠詞の使用方法に関する訂正理由、文脈に基づく例文の3つの情報を提示する手法を提案した。評価実験では、冠詞誤り訂正モデルの冠詞の分類性能の評価と、根拠提示の有効性を検証するための評価を行い、自動訂正時に提案した訂正根拠を提示することの学習者の冠詞選択能力への有効性を示した。

5.2 今後の方針

本稿では誤り訂正根拠の提示手法の初期の実験冠詞の使用方法に関する誤りについて焦点を当てて実験を行ったが、この手法は前置詞誤りなど分類問題に定式化できる他の文法誤り訂正に対しても適用が可能である。今後の方針としては前置詞誤り訂正などの他の誤りへの本手法を適用したときの影響や課題を分析することを考えている。さらに、分類モデルの学習に用いる素性を根拠情報の提示という側面を考慮して精巧に設計することや、訂正根拠の提示が学習者に与える学習効果の検証をすることのできる長期の評価実験を行う必要がある。また、学習者の母語によって英文の誤り方の違いがあるため、母語に合わせた根拠の提示方法や学習データの利用を検討する。

謝辞

本研究を進めるにあたって，多くの方にご協力をいただきました．ここに，心より感謝の意を表します．

乾健太郎教授には，お忙しい中，研究活動全般にわたり，終始手厚いご指導，ご助言をいただきました．心より感謝を申し上げます．ご多忙の中，審査委員をお引受けくださった，加藤寧教授，伊藤彰則教授に深く感謝致します．本研究を進めるにあたり，適切なご助言をくださいました，岡崎直観准教授，渡邊陽太郎助教，松林優一郎研究特任助教，水野淳太研究員，Eric Nichols 研究員に感謝致します．また，研究活動および大学生活を暖かく支えてくださいました，八巻智子秘書に感謝致します．

最後になりましたが，研究室での生活から研究に関する議論まで，多くの面で研究活動を支えてくださった乾・岡崎研究室の皆様我心より感謝致します．

参考文献

- [1] Diane Nicholls. The cambridge learner corpus - error coding and analysis for lexicography and elt. In *Corpus Linguistics 2003*, pp. 572–581, 2003.
- [2] Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel R. Tetreault. *Automated Grammatical Error Detection for Language Learners*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2010.
- [3] Daniel Dahlmeier, Hwee Tou Ng, and Eric Jun Feng Ng. NUS at the HOO 2012 Shared Task. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pp. 216–224, 2012.
- [4] Robert Dale and Adam Kilgarrieff. Helping our own: Text massaging for computational linguistics as a new shared task. In *Proceedings of the 6th International Natural Language Generation Conference*, pp. 261–265, 2010.
- [5] M. Chodorow, M. Gamon, and J. Tetreault. The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing*, Vol. 27, No. 3, pp. 419–436, 2010.
- [6] Na-Rae Han, Martin Chodorow, and Claudia Leacock. Detecting errors in English article usage by non-native speakers. *Natural Language Engineering*, Vol. 12, pp. 115–129, 2006.
- [7] Guido Minnen, Francis Bond, and Ann Copestake. Memory-based learning for article generation. In *Proceedings of the 2nd workshop on Learning language in logic and the 4th conference on Computational natural language learning (CoNLL '00)*, pp. 43–48, 2000.
- [8] Michael Gamon, Jianfeng Gao, Chris Brockett, Alexandre Klementiev, William B. Dolan, Dmitriy Belenko, and Lucy Vanderwende. Using contextual speller techniques and language modeling for ESL error correction.

In *Proceedings of the Third International Joint Conference on Natural Language Processing (IJCNLP 2008)*, pp. 449–456, 2008.

- [9] Michael Gamon. Using mostly native data to correct errors in learners’ writing: a meta-classifier approach. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT ’10)*, pp. 163–171, 2010.
- [10] Ryo Nagata and Kazuhide Nakatani. Evaluating performance of grammatical error detection to maximize learning effect. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING ’10): Posters*, pp. 894–900, 2010.
- [11] Emi Izumi, Kiyotaka Uchimoto, Toyomi Saiga, Thepchai Supnithi, and Hitoshi Isahara. Automatic error detection in the japanese learners’ English spoken data. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 2*, pp. 145–148, 2003.
- [12] John Lee. Automatic article restoration. In *HLT-NAACL 2004: Student Research Workshop*, pp. 31–36, 2004.
- [13] Martin Chodorow, Joel Tetreault, and Na-Rae Han. Detection of grammatical errors involving prepositions. In *Proceedings of the Fourth ACL-SIGSEM Workshop on Prepositions*, pp. 25–30, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [14] Rachele De Felice and Stephen G. Pulman. A classifier-based approach to preposition and determiner error correction in L2 English. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pp. 169–176, 2008.
- [15] Joel R Tetreault and Martin Chodorow. The ups and downs of preposition error detection in ESL writing. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling ’08)*, pp. 865–872, 2008.

- [16] Na-Rae Han, Joel R. Tetreault, Soo-Hwa Lee, and Jin-Young Ha. Using an error-annotated learner corpus to develop an ESL/EFL error correction system. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC '10)*, pp. 763–770, 2010.
- [17] Michael Gamon. High-order sequence modeling for language learner error detection. *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 180–189, 2011.
- [18] Daniel Dahlmeier and Hwee Tou Ng. Grammatical error correction with alternating structure optimization. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 915–923, 2011.
- [19] Chris Brockett, William B. Dolan, and Michael Gamon. Correcting ESL errors using phrasal SMT techniques. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pp. 249–256, 2006.
- [20] Matthieu Hermet and Alain Désilets. Using first and second language models to correct preposition errors in second language authoring. In *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 64–72, 2009.
- [21] Xing Yi, Jianfeng Gao, and William B. Dolan. A web-based English proofing system for English as a second language users. In *Proceedings of the Third International Joint Conference on Natural Language Processing (IJCNLP 2008)*, pp. 619–624, 2008.
- [22] Alla Rozovskaya and Dan Roth. Algorithm selection and model adaptation for ESL correction tasks. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 924–933, 2011.

- [23] Robert Dale, Ilya Anisimoff, and George Narroway. HOO 2012: A report on the preposition and determiner error correction shared task. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pp. 54–62, June 2012.
- [24] Edward Loper and Steven Bird. NLTK: the Natural Language Toolkit. In *Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics - Volume 1*, pp. 63–70, 2002.
- [25] Yoshimasa Tsuruoka, Yuka Tateishi, Jin-Dong Kim, Tomoko Ohta, John McNaught, Sophia Ananiadou, and Jun’ichi Tsujii. Developing a robust Part-of-Speech tagger for biomedical text. In *Advances in Informatics*, Vol. 3746, pp. 382–392, 2005.
- [26] Steven Bird, Robert Dale, Bonnie Dorr, Bryan Gibson, Mark Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir Radev, and Yee Fan Tan. The ACL Anthology Reference Corpus: A reference dataset for bibliographic research in computational linguistics. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC ’08)*, 2008.
- [27] John Lee and Stephanie Seneff. Automatic grammar correction for second-language learners. In *Proceedings of the Ninth International Conference on Spoken Language Processing (Interspeech)*, pp. 1978–1981, 2006.
- [28] Ryo Nagata, Tatsuya Iguchi, Kenta Wakidera, Fumito Masui, Atsuo Kawai, and Naoki Isu. Recognizing article errors using prepositional information. *Systems and Computers in Japan*, Vol. 37, No. 12, pp. 17–26, 2006.
- [29] Jenine Turner and Eugene Charniak. Language modeling for determiner selection. In *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 177–180, 2007.

発表文献一覧

国際会議論文

- Jiro Umezawa, Junta Mizuno, Naoaki Okazaki, and Kentaro Inui. Evidence in Automatic Error Correction Improves Learners' English Skill. In *Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2013)*, (to appear).

国内会議・研究会論文

- 梅澤次郎, 水野淳太, 岡崎直観, 乾健太郎. 冠詞誤り訂正時における訂正根拠の提示. 言語処理学会第 18 回年次大会論文集, pp.353-356, 2012.